

# Web Speech API を利用した韓国語音声認識クイズの開発

須賀井 義教

## 1. はじめに

本稿は、Web Speech API<sup>1</sup> を利用した、ウェブブラウザで利用可能な音声認識による韓国語単語クイズ（以下「音声認識クイズ」と呼ぶ）の開発とその実践例について述べるものである。

近年、デスクトップパソコンだけでなく、タブレットやスマートフォンなどの利用が広がるなかで、コンピュータを利用した外国語学習（CALL）が存在感を増してきている。特に英語教育においては、いわゆる CALL 教室を利用して、さまざまなソフトウェア、アプリケーションを駆使した教育が行われている<sup>2</sup>。また、授業や教室における活用だけでなく、授業外の学習をサポートするような教材も開発され、時と場所を選ばず、自分のペースで学習できるという CALL のメリットが大いに活かされているものとみられる。

筆者もこれまで、自らが運用するホームページを通じて、様々な補助教材やツールを提供してきた<sup>3</sup>。公開している補助教材は全てブラウザを通じて利用するものであり、ネットワーク通信が可能であればどこでも利用できる。特に、授業で用いているテキストの内容に即した「単語クイズ」は、月間 700 ほどのページビュー<sup>4</sup>があり、授業外の学習において多く利用されていると見られる。

ウェブブラウザだけで利用できる CALL 教材の利点は、何と云っても学習者の利用環境を選ばないことであろう。専用のソフトを必要とする教材の場合、それがインストールされていなければ利用することができず、開発の際にもデバイスごと、あるいはオペレーティングシステムごとにプログラムを作成しなければならない。一方、ウェブブラウザはほぼ全てのデバイスに搭載されており、学習者も手元の機器ですぐに学習を始めることができるし、開発者も一度のプログラミングで済む。

こうした利点を念頭に、筆者はブラウザで利用できる教材を作成してきたが、今回、Web Speech API について知る機会があり、これを応用して音声認識クイズを開発した。本稿ではその概要と、授業における実践について報告する。

## 2. Web Speech API について

注 1 で述べたように、Web Speech API はブラウザで音声認識と音声合成の機能を利用できるようにするものである。Google によるデモ（URL は注 1 を参照）は、音声を認識

してその内容を文字化し、クリップボードにコピーしてくれるものだが、英語や日本語だけでなく、さまざまな言語に対応していることが分かる。

無償で利用できる音声認識エンジンとしては他に Julius<sup>5</sup> があるが、サーバーを設置する必要がある、外国語への対応が困難であるなどといった点で、利用の際の敷居が高い。一方、本稿で利用した Web Speech API は無償で使用できるだけでなく、JavaScript で簡単にプログラミングでき、さまざまに活用することができる。例えば、音声の認識と合成の両者を利用して、英語の発音学習ができるページ<sup>6</sup> もあり、外国語教育向けの用途にも利用可能であろう。

なお、Web Speech API は現在のところ Google Chrome と Apple 社の Safari で部分的に実装されている。Chrome ではスマートフォン向けのバージョンと iOS 版を除き、音声認識と音声合成のいずれも利用できるが、Safari では音声合成のみサポートしている。よって本稿では、Chrome での利用を前提として開発を進めた<sup>7</sup>。なお、利用の際にはマイクなどの音声入力デバイスとネットワーク接続が必要である。また、Web Speech API と jQuery<sup>8</sup> を利用しているため、ブラウザ側で JavaScript をオンにしておかなければならない（通常はオンになっている）。

### 3. 音声認識クイズの概要

#### 3.1 クイズの内容

本稿で紹介する音声認識クイズは、示された日本語に対し、対応する韓国語の単語を学習者がきちんと覚えているか、またその語を正確に発音できているか、自分でチェックするための補助教材として開発したものである。現在のところ、その内容は筆者が担当している「ハングル」能力検定試験<sup>9</sup> 対策の講義で使用するために作成したテキストに基づいている。テキストは初級段階の 5 級と 4 級向けの内容で、それぞれの課で扱われる語彙は試験の出題基準に合わせてある。

なお、この音声認識クイズでは、各課の語彙から任意に 10～15 個を選び、出題することにした。各課の新出語彙は 20～30 程度であるが、後述するように、音声認識クイズにタイムトライアルという性格を持たせており、出題数が多いとやや間延びしてしまうおそれがある。そのため、クイズでは全ての新出語彙を提示せず、出題数をやや少なめに制限した。

テキストは私家版で未公開のため、出題語彙のリストのみ、別途公開することとした<sup>10</sup>。現在は 5 級向けの問題を一部の課に限定して公開しているが、今後問題を追加し、更新していく予定である。

### 3.2 音声認識クイズの利用方法

音声認識クイズ本体は既にインターネット上で公開している。 <http://porocise.sakura.ne.jp/korean/kentei/stt/> から利用可能である。初期画面は図1の通り。画面上部には開始や停止など、主にコントロール関連の部品を配置してある。画面最下部の灰色部分には、エラーの際のメッセージやクイズの終了などを表示するようになっている。



図1 音声認識クイズの初期画面

画面上部、左の部分にあるドロップダウンリストから任意の課を選択し、「Start」ボタンを押すと、Chrome からマイクへのアクセスを確認する警告が表示される（図2）。

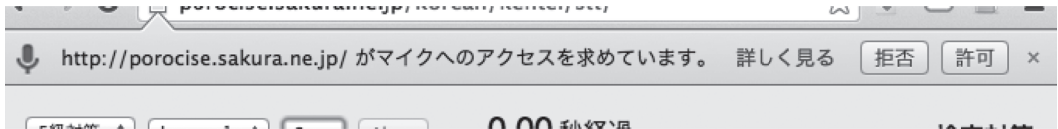


図2 マイクへのアクセス確認画面

ここで「許可」を押すと音声認識が開始され、問題が提示されるとともに、経過時間のカウントがスタートする<sup>11</sup>。画面上部に日本語が提示されるので、それに対応する韓国語の単語をマイクに向かって発音すると、認識された結果が灰色の文字で画面に表示される(図3)。文字が灰色の間はまだ音声認識が続いており、発音すればさらに認識内容が追加される。発音をやめ、音声入力が途切れると、そこで認識内容が確定され、文字色が黒となる。

確定した認識内容が正解であれば、赤い○が表示されて次の問題が即座に出題されるが(図4)、不正解の場合、続けて発音しなければ文字が消え、再度発音しなければならない。



図3 音声認識中の画面表示(文字色：灰色)



図4 認識内容確定、正解の場合(文字色：黒)

選択した課の全ての問題に正解すると、クイズは終了となり、経過時間のカウントアップがストップする(図5)。認識開始から全問正解まで、目安ではあるが、およそ何秒かかったか分かる仕組みとなっている。



図 5 クイズの終了画面

クイズを途中でやめるには「Abort」ボタンを押せばよい。その後「Start」ボタンを押した場合は、問題が途中から再開されるのではなく、選択した課の問題が新たに開始される。

#### 4. 授業における実践の例

音声認識クイズはインターネット上で公開しているため、自習の際に用いることもできるが、学習者が一人で利用する際、発音に問題があって先に進むことができないことがあると思われる。自分の発音がどのように誤っているのか、どのように矯正すればよいのか分からず、先に進めないようなケースである。その点では、教師のいる教室で利用する方が、発音の指導も行うことができ、より効果的であるとも考えられる。

以下では、実際に筆者が担当する講義での実践を紹介する。また、教室での利用を想定した機能についても述べる。

#### 4.1 実践の内容

2014年度前期に筆者が担当した講義のうち、受講者数が比較的少ない二つのクラスにて、音声認識クイズを数回利用した。いずれのクラスも初級であるが、一つは2014年4月から韓国語の学習を開始した学習者を含んでおり、もう一つは学習を始めて少なくとも1年は経過した学習者のクラスである。仮に前者のクラスを初級の「前半クラス」、後者を「後半クラス」と呼ぶこととする。

前述の通り、音声認識クイズでは全問正解に至るまでのおおよその経過時間が計測され、その時間がほぼリアルタイムで表示される<sup>12</sup>。この機能を利用して、受講者どうし、タイムを競わせることにした。

なお、今回の実践では各受講者を一人ずつパソコンの前に立たせて発音させ、クイズの画面をプロジェクターで映し出すことにした。この実践では受講者が少ないこともあり、また人前で発音するという緊張感も持たせるため、こうした方法をとった。もちろん、CALL教室などを利用して、各自がそれぞれの端末でクイズを行い、最終的にベストタイムを申告する、という方法も可能である。

プロジェクターで画面を投影する場合、認識された文字が小さく表示されると見にくいいため、文字数に関係なく、フォントサイズを自動的に調整し、文字全体がブラウザの横幅いっぱいに表示されるようにしてある<sup>13</sup>。

教室の設備として、パソコンおよびマイク、プロジェクター、ネットワーク接続が必要である。パソコンにはGoogle Chromeブラウザがインストールされている必要がある。

実施に先立って、クイズを行う課を一つ選び、単語の発音を受講生とともに確認する。時間を与え、日本語を見て韓国語の単語を発音できるようにさせる。

ある程度発音に習熟し、単語を覚えたことが確認できたら、回答する順番を決め、一人ずつパソコンの前に立たせてクイズを行う。単語の暗記に不安がある場合には、単語のリストを見ながら行ってもよい。また、途中うまく認識されない、あるいは受講生がうまく発音できない場合には、適宜発音のポイントをヒントとして与える。何度やってもうまくいかない場合には、教師が代わりに発音し、次の問題に移るのがよいと思われる。

各自全問終了の際には、最後に表示される経過時間を黒板などに書き出しておく。全員の回答が終わった時点で、タイムを比較し、最も短かった受講生が勝者となる。

実際に行なってみたところ、10問の問題で最短30秒～40秒程度かかり、最長で150秒程度という結果であった。

#### 4.2 受講生の反応

最終回の講義で、出席者にコメントを書いてもらった。前半クラスでは感想、良かった

点、悪かった点および要望についてコメントしてもらい、後半クラスでは記述の時間が不足したため、感想のみ書いてもらった。以下、全員の意見を載せる（原文のまま）。

#### 4.2.1 感想

- ・自分の発音が反映されて、面白かった。単語を覚えるだけでなく、自ら発音してみると、さらに記憶に刻まれるように感じた。（前半クラス・女）
- ・発音の練習にもなるし、何人かの人数ですると楽しみながらできておもしろかった。（前半クラス・女）
- ・認識してくれたときは、うれしかった。（前半クラス・男）
- ・時間を忘れるくらい楽しかった。（前半クラス・男）
- ・音声認識のやつは、自分が思っているのとは違う音が出てることがわかってよかったです。スとヌのちがいがむずかしかったです。（後半クラス・女）
- ・舌がとてもむずかしかった。何回か練習したら、本番1発でできたのでとてもうれしかった。タイムとかあって、みんなでわいわいできたので楽しく学べてよかった。（後半クラス・女）
- ・単語覚えれた。おもしろかった。笑えた。（後半クラス・女）
- ・音声を認識する感度が思ったよりも良かったので驚いた。発音練習になるので良かった。（後半クラス・男）
- ・音声など、他ではあまりしていないことだったので、少人数ならではで、とても楽しかったです。（後半クラス・女）

#### 4.2.2 良かった点

- ・声がすぐに反映される。（前半クラス・女）
- ・時間もあって競争しながらできるから頑張ろーってなる。（前半クラス・女）
- ・時間がカウントされているから、緊張しつつも、クラスメートと競争して、ゲーム気分でやれるのはいいと思う。（前半クラス・男）
- ・自分が正しい発音で話せているか分かる。とっさの語彙力が問われる。失敗した後に成功した達成感が味わえた。（前半クラス・男）

#### 4.2.3 悪かった点, 要望

- ・発音が認識しにくい単語があるようで、何度も繰り返して言うのが腹立たしい。何度か間違えたら、その単語をとばすなどの機能があれば良いなと思った。（前半クラス・女）
- ・まちがえたのはあってる発音を言ってくれたら分かりやすいかなと思いました。（前半ク

ラス・女)

- ・ ランキングやタスクなどの機能を追加したら、もっといいソフトになると思う。(前半クラス・男)
- ・ 雑音がたまに入る。「ㄥ」と「o」の区別があまりされなかった。(前半クラス・男)

以上のように、概ね好意的な評価を得た。特にクラスメートと競争するという点に面白さを覚えたように見受けられる。一人での作業や自習ではなく、クラスで行うという点の良さであると言えよう。

また、要望のうち、問題をスキップする機能については今後実装を検討していきたい。例えば、認識を3度しても正解しない場合に自動的に次の問題に移ったり、あるいはボタンを押すとその問題をスキップしたり、など方法はいくつか考えられる。経過時間を競うということを考えると、スキップ1回につき最終タイムに10秒追加、などといったペナルティーも必要かもしれない。

他の要望のうち、正解の発音を音声として提示してほしいという指摘も重要である。特に自習者に対して助けになるものと思われる。

今回の簡単なアンケートではとりあえず好意的な意見が多かったが、実際にこの音声認識クイズを利用することでどのような学習効果が見られるか、検証する必要もあろう。発音教育においてどのような効果を挙げることができるか、課題の一つとして今後検討していきたい。

## 5. 利用に際しての注意

なお、音声認識クイズの利用に際して、いくつかの注意点がある。

### 5.1 音声入力に関する注意

まず、マイクからの音声入力を全て拾ってしまうため、言いよどみや言い直しの音声、周囲の雑音なども入力として受け取り、認識してしまう。特に「あ」「えー」などといった言いよどみが入力として認識されてしまうことが多いように見受けられる。Web Speech APIで認識が終了した際に発生するresultイベントを検知して、作成したプログラムの側で正解の判定を行っているが、「認識終了」のタイミングと音声の切れ目とがうまく合っていないように思われる。今後も検討をしていく予定である。

また、上記とも関連するが、認識内容の確定までにかかる時間が一定でなく、文字の色が灰色(認識途中)から黒色(認識内容確定)に変わるまで、やや待たされる場合がある。この時、認識されなかったものと誤解し、続けて発音してしまうと、認識内容がさら



に追加されてしまう。

これらの問題は、マイクの感度を変えるなど、コンピュータ側の設定により解決されるかもしれない。一時的に音声をミュートするスイッチがついたマイクなどもあるため、こうした機器の利用を検討してみたい。また、認識内容の確定までにかかる時間はネットワーク環境などにも左右されるため、高速なネットワークの利用が望ましいと考えられる。単語を発声したら確定されるまで黙っている、というように利用者側も注意する必要がある。

## 5.2 音声認識に関する注意

以上は音声の入力に関わる注意点であるが、他にも Web Speech API の音声認識に関わる注意点がある。ここでは認識の精度などについて述べておく。

実際に利用してみればすぐ分かると思われるが、どれほど正確な発音を心がけ、細心の注意をもって発音しても、思い通りに認識されない語がある。また反対に、かなりルーズに発音してもうまく認識されてしまう語もある。

こうした Web Speech API の認識精度について、ユーザーあるいは作成者がプログラムの側で操作したり、調整したりすることはできない。何度やってもうまく認識されず、正解にたどり着けなければ、学習者のやる気を削ぐ結果となってしまうため、現在は筆者が実際に発音してチェックの上、確実に認識される語だけを選んで出題してある。第4節で紹介したように教室などでクイズを行う場合には、教師が事前に出題語彙の全てをチェックし、適切に発音の指導ができるよう、準備をしておくべきであろう。

また、本稿で紹介した音声認識クイズでは、Web Speech API からの認識結果が一つしか採択されない。そのため同音異義語がある場合には、出題内容と認識結果が一致しない可能性がある。例えば「口」(입)と「葉」(잎)とは綴りが異なるが、単独での発音はいずれも [입] である。実際に認識させてみると、API からは「口」(입)という結果しか返されない。この音声認識クイズでは、認識結果として API から返された文字列と、正解の文字列とが一致するか、作成したプログラムの側で判定するという単純な仕組みのため、同音異義語に対応するのは困難である<sup>14</sup>。

文字列の一致で判定するという仕組みが引き起こすその他の問題として、字種と分かち書きがある。字種について、例えば [이] と発音すると、「この」という冠形詞や「歯」という名詞ではなく、数詞の「2」(이)として認識される。この場合、API からは認識結果としてアラビア数字の「2」が返されるため、正解をハングル表記「이」としておくと、不一致となり正解と判定されない。数詞は漢数詞、固有数詞を問わず、基本的にアラビア数字が認識結果として返されるようである。そのため、例えば「7時」の正解を「일곱

시]のようなハングル表記ではなく、「7 시」のようにアラビア数字で表記しておくなど、数詞が含まれる問題の作成には工夫が必要と思われる。

また分かち書きについては、文を発音した場合、正書法に基づいた分かち書きがされていない結果が返されることがある。教育上、適切な分かち書きで表示されるのが望ましいが、この部分も API から返される結果をそのまま利用するしかなく、プログラム側で調整ができない。現在の出題内容のうち、分かち書きが問題になるケースは含まれていないが、念のため、認識結果の文字列から全てのスペースを削除して、正解と一致するか判定するようにしている。

ここまで、音声認識クイズ利用の際の注意点をいくつか述べてきた。利用の際だけでなく、出題あるいは Web Speech API を用いたプログラム作成の際の注意点とも言える。今後の参考とされたい。

## 6. おわりに

ここまで、Web Speech API を利用して筆者が開発した、音声認識クイズの概要とその利用実践例、さらに利用の際の注意点を述べてきた。

音声認識クイズの最大の特徴は音声によって解答を入力するという点であり、単に正解をクリックするといった四択クイズに比べ、学習者とコンピュータの間のインタラクションがさらに増え、学習者の自律的な学習を促進するという点にも合致している（吉田晴世ほか編著 2008:20-21）。また、教室での発音練習とは異なり、コンピュータがあれば一人で練習することが可能であるため、他の学習者の前で発音することへの不安感を取り除くことができる。国際交流基金（2009:145-146）では発音練習における教師の役割として、学習者が目標の発音をイメージできるようになることを手助けする、またそれを元に学習者が自分の発音を自己モニターできるようになることを手助けする、といったことに加え、学習者がいつでも不安なく練習できるような環境や雰囲気を作っていくことも挙げている。本稿で紹介した音声認識クイズは、まさにこうした教師の役割、特に練習環境作りを肩代わりすることが可能であろう。

ただ、プログラム自体に改善の余地があるのも確かであり、実践の中で受講生から寄せられた要望などについて、今後検討し実装していきたいと考える。特に自習者向けの機能、例えば単語ごとに発音のポイントを表示させたり、激音や濃音などといった発音の特徴ごとに練習を行なえるようにしたり、といった機能を追加する必要があるだろう。発音を矯正するという観点からは、モデル音声の提示もやはり必要と考える。Web Speech API の音声合成を利用すれば、別途 MP3 などの音声ファイルを準備せずに、実装が可能

である。

なお、本稿で紹介した音声認識クイズは、あくまで音声認識の活用例の一つに過ぎない。クイズ以外にも、シャドーイングの音声を文字化して、元の文章と比較するなどといった、音声認識の機能を利用した教材開発が可能であると考え。また、Web Speech APIは韓国語だけでなく様々な言語に対応しているので、他言語にも応用できる。一つの画面上で学習したい言語を選択し、同じ出題形式で音声入力によるクイズを解くことができる、といった多言語教材も考えられる。本稿での取り組みが、今後の参考となれば幸いである。

## 注

- 1 Web Speech APIは、W3C (World Wide Web Consortium) のコミュニティグループである Speech API Community Group により策定されている JavaScript API で、各種言語に対応した音声認識 (Speech Recognition) と音声合成 (Speech Synthesis) の機能をウェブページに提供するものである。W3C 標準ではないため、実装はブラウザにより異なる。仕様は <https://dvcs.w3.org/hg/speech-api/raw-file/tip/speechapi.html> を参照のこと。Google によるデモ (<https://www.google.com/intl/ja/chrome/demos/speech.html>) も参考にされたい。
- 2 具体的な事例として、本学総合社会学部の英語科目におけるオンライン英文エッセイ添削サービス「Criterion」の導入、実践などが挙げられる。西村香奈絵 (2013) 参照。
- 3 <http://porocise.sakura.ne.jp/wiki/> を参照。「ハングル」能力検定試験向けの作文練習や、李潤玉ほか (2011) などに対応した単語クイズ、シャドーイング用の素材、フラッシュカード作成ツールなどがある。
- 4 サイト内に設置したアクセス解析のデータ (2014年6月22日から7月22日までの1ヶ月間) による。
- 5 <http://julius.sourceforge.jp/>
- 6 「Vois.io Pronunciation」 (<https://vois.io/ja/pronunciation>)。以前にはアクセスできたが、本稿執筆時点では接続ができなかった。
- 7 本稿で動作を確認したのは、Windows 7とMac OS X (Mountain LionおよびMavericks) 上のChrome (バージョン37)、Android (4.4 Kit Cut) 上のタブレット向けChrome (バージョン37) の組み合わせである。ただし、Androidタブレットではデバイス上の音声認識言語を韓国語にしておく必要がある上、画面のレイアウトが

崩れるなど利用にあまり向いていない。

- 8 JavaScript ライブラリの一つ。本稿ではバージョン 1.9.1 を使用した。 <http://jquery.com/>
- 9 ハングル能力検定協会が実施する、日本語母語話者向けの韓国語検定試験である。1993 年から実施され、現在は年 2 回、6 月と 11 月に実施されている。
- 10 [http://porocise.sakura.ne.jp/wiki/korean/kentei/word\\_list](http://porocise.sakura.ne.jp/wiki/korean/kentei/word_list) で公開中。
- 11 なお、「拒否」を押すとマイクからの入力を受け付けなくなり、クイズは開始されない。その場合にはマイクへのアクセスが拒否された旨、メッセージが表示されるようになっている。
- 12 「おおよその経過時間」「ほぼリアルタイム」としているのは、プログラムの都合上、正確な時間を測ることができていない可能性があるためである。例えば「Abort」ボタンを押した際、即座に計測がストップしない。この点については今後プログラムを改善していくが、経過時間はあくまで目安であることを了解されたい。
- 13 具体的には、jQuery プラグインの一つである slabText (<https://github.com/freqDec/slabText/>) を用いている。
- 14 実際には、Web Speech API では複数の認識候補について、候補のリストとその信頼度スコアを取得することができる（仕様 5.1.5 節を参照）。このリストとスコアを利用して、同音異義語に対応させることが可能かもしれない。今後の課題としておく。

### 参考文献

- 李潤玉ほか（2011）『改訂版 韓国語の世界へ 入門編』、朝日出版社
- 国際交流基金（2009）『国際交流基金 日本語教育シリーズ 第2巻 音声を教える』、ひつじ書房
- 西村香奈絵（2013）「英語エッセイライティングクラスにおいて学生のエッセイに見られた変化—総合社会学部1年生のある中位クラスにおける事例—」、『近畿大学総合社会学部紀要』第3巻第1号、近畿大学総合社会学部
- 吉田晴世ほか編著・CIEC 外国語教育研究会著（2008）『ICT を活用した外国語教育』、東京電機大学出版局